USING DENTAL METRICS TO PREDICT GENDER

GEEKLURN®
Right Place for Technology Learning

PROJECT DETAIL

Project Title	Using Dental Metrics to Predict Gender
Tech Stack	Python, Machine Learning
Domain	Healthcare
Project Difficulty level	Rookie/ Basic
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel

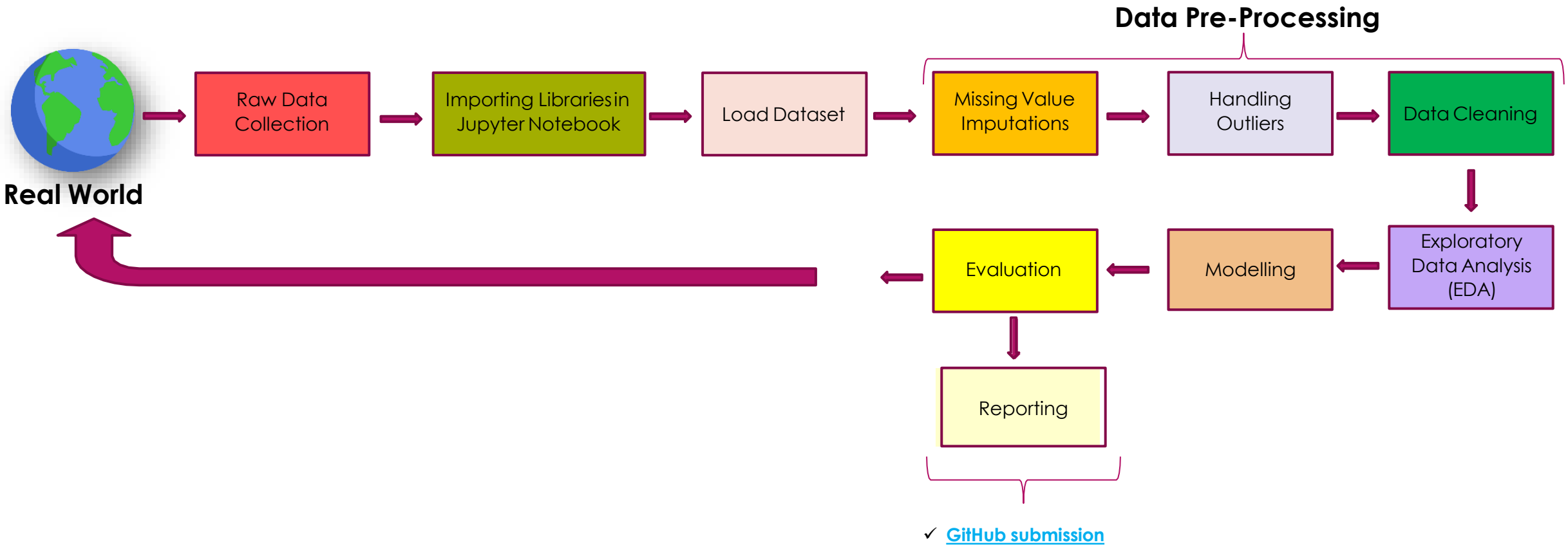
OBJECTIVE

- The goal of this project is to analyse the data and predict, based on a combination of dental features that describes the Gender of the person

BACKGROUND & SCOPE

- Forensic medicine is an interesting area of study. Forensic dentistry is a branch of forensic medicine.
- During natural calamities or due to some other reasons, many times, it will not be possible to find out the gender of the deceased person.
- In such cases, certain measurements of the tooth will be taken (as bones and teeth do not decay easily) and gender will be determined.

ARCHITECTURE



DATASET INFORMATION

Age: The person's age in years

Gender: The person's sex (male, female)← Target Variable

SampleID and SL No. : The sampleID & SL No. represents individuals unique ID

Inter-canine distance intraoral , inter-canine distance casts, right canine cast, left canine cast, etc. These features represent the measurement of the oral teeth ← Our Independent Variables

Step-by-Step Approach to follow:

Step 1: Raw data collection : Click the hyperlink to download the dataset - Hyperlink

Step 2: Importing the necessary packages in JupyterNotebook/ Any IDE

Note: For JupyterNotebook Installation kindly follow the documentation

Packages involved:-

import pandas as pd #Used to load the dataset

import Numpy as np #Used to perform mathematical operations

import matplotlib.pyplot as plt }

import seaborn as sns } #Used to visualize the data

Step 3: Import the dataset using pandas

Variable_name = pd.read_csv("Dentistry.csv")

Step 4: Data Preprocessing

i) Identify and handle missing values

ii) Encoding categorical data

i.e from sklearn.preprocessing import LabelEncoder

Step-by-Step Approach to follow cont.

iii) Split independent and dependent variables i.e. X and Y
iv) Normalize the X variable
from sklearn.preprocessing import Normalizer #all the values will fall in the range [0,1] or sometimes[-1 , +1]

Step 5: Exploratory Data Analysis
i) You need to check the correlation of the data using Heatmap between X-to-X features and X-to-Y features to understand the relationship and collinearity issues between the features. (seaborn library)

Step 6: Model Building
i) Drop the unwanted independent variables which you see not important for model building.
ii) Drop the independent features which are highly correlated to each other
iii) Split the Data into Train and Test set
from sklearn.preprocessing import train_test_split
iv) Use Logistic Regression, Decision Tree classifier, Random Forest classifier and XGBoost classifier.

Step-by-Step Approach to follow cont.

Step 7: Evaluation

i) You need to evaluate the model based on the models evaluation metrics i.e. Confusion matrix(Accuracy), ROC curve and AUC curve to check model accuracy and plot them

ii) Step 8: Goto GitHub Link and create a folder in your name and upload the code file there as submission process

iii) **Note:** Unless you do not create your own GitHub account you will not be able to access / submit the project. Create a GitHub Account with the email id registered in GeekLurn.

THANK YOU